Original article

Retrospective study to validate and develop an objective assessment of the degree of eye redness using video topographers

Wolfgang Sickenberger[1,3], Sebastian Marx[2,3]

[1] Prof., MS Optom (USA), Dipl.-Ing. (UAS) AO · [2] PhD, M.Sc., Dipl.-Ing. (UAS) AO, FIACLE · [3] JENVIS Research c/o Ernst Abbe University of Applied Sciences Jena, Jena, Germany

Abstract

Purpose. The purpose of this exploratory study was to evaluate and validate the objective determination of bulbar, limbal and total conjunctival redness with the K5M video topographer (Oculus, Germany), the measurement software (R-Scan) and the JENVIS grading scale in comparison to subjectively recorded redness levels.

Material and Methods. The conjunctival redness level was documented three consecutive times in a total of 75 subjects (150 eyes). The average age was 33.9 ± 15.7 years. The survey took place at the Centre for Contact Lens Research Waterloo, Canada and at the Helios hospital Erfurt, Germany. From the survey, 25 images were selected and entered into a website for subjective grading. On the website, the images were presented in a randomized order six times (3× limbal & 3× bulbar redness), so that each of the 20 investigators graded a total of 150 images. The reference images of the redness grades were based on the JENVIS grading scale. The subjective results were compared using the five analyzed grades with the objective findings by the R-Scan of the K5M. The results were analyzed using descriptive statistics.

Results. With regard to the images obtained using the software, the mean grade of findings was 1.3 ± 0.3, with 24 % of the results being assessed as grade 0, 48 % as grade 1, 14 % as grade 2 and grade 3. The subjective findings resulted in a mean grade of 1.5 ± 0.3. 17 % were assessed as grade 0, 45 % as grade 1, 19 % as grade 2 and 5 % as grade 3. The results of the subjective and objective classification of temporal and nasal bulbar redness are not statistically different. In contrast, the results of temporal and nasal limbal redness are statistically different, with the subjective findings within grade 1 being statistically significantly higher than those of the objective findings. The mean dispersion of the objective classifications is 11 %, that of the subjective 31 %.

Conclusion. The software (R-Scan, K5M) primarily evaluates the number and redness of the conjunctival vessels, whereas the subjective evaluation is based on the overall redness of the eye as perceived by the examiner. As the degree of severity increases, the spread of the results increases with objective classification, but decreases with subjective classification, with the latter showing an almost 3-times greater spread on average. This means that automated conjunctival classifications are more accurate than subjective classifications for lower degrees of severity and comparable for higher degrees of severity.

Keywords
classification, conjunctival redness

Introduction

The conjunctiva of the eye consists of a large number of blood vessels and fine capillaries. When irritated, the vessels dilate and supply the tissue with more blood; this is referred to as conjunctival injection or hyperaemia. Oxygen exchange is stimulated, and an increased immune response occurs. Tarsal, limbal and bulbar redness of the eye are therefore caused by hyperaemia of the blood vessels in the conjunctiva and/or episclera.[1] This is known as conjunctival or ciliary injection.[2] The causes in such a case are not always specific.[1,3] Possible causes for these reactions are diverse and can, for example, be attributed to environmental conditions, contact-lens wear, allergies or various (eye) diseases. Other factors may be metabolic, chemical, toxic and mechanical or the effects of keratoconjunctivitis sicca.[4,5] To evaluate eye redness, a variety of grading scales based on photographs, as well as drawn or computer-generated images, have been developed.[4] Examples of grading scales used worldwide are the Cornea and Contact Lens Research Unit (CCLRU) grading scales (Brien Holden Vision Institute),[6] Efron grading scales[4] and JENVIS grading scales.[7] As early as 1889, August Müller reported the distinction between bulbar and limbal redness, which is used as standard in most grading scales today.[8] Murphy et al. described a mean bulbar redness grade of 1.9, determined with the non-validated CCLRU classification key as physiologically normal (mean age of the test subjects: 29 years).[9] "A bulbar redness of greater than grade 2.6 may be considered abnormal, and a grade change in bulbar redness of > or = 0.4 may be significant".[9] In an older test group with a mean age of 45 years, mean bulbar and limbal redness levels above 2.8 and 2.5, respectively were considered as symptomatic.[9,10] One problem with printed grading scales is their lower precision due to subjective judgement.[11] Furthermore, assessment with different grading scales for the same finding can lead to different results.[12,13]

Aim of the study

The aim of this exploratory study was to evaluate and validate an objective determination of the degree of eye redness of the anterior segment of the eye that was developed in 2012.

Material and Methods

Selecting the images

The data analysis is based on images from 75 test subjects (38 female and 37 male) and accordingly 150 eyes. The average age of the subjects was 33.9 ± 15.7 years. The images were collected between June and December 2012 at the Centre for Contact Lens Research (University of Waterloo, Canada) and at the Helios Hospital in Erfurt (Germany) by one examiner per centre. 25 images were selected from a pool of 955. The selection of the images for objective and subjective classification was based on the criteria of image sharpness of different conjunctival conditions and heterogeneity in terms of severity, pathology and ethnic origin of the test subject (Table 1). The eyes were categorised as Caucasian, Oriental, Asian and African. Due to a reduced number of findings with ocular pathological background, different original images of the same pathology were shown several times. These included pathologies such as blepharitis, glaucoma, stenosis or insufficiency of the lacrimal ducts.

Examiner collective

A total of 135 experts were contacted, 35 of whom wanted to support the study. Inclusion criteria for examiners were, for example, a qualification in optometry or ophthalmology with completed training, as well as more than three years of professional experience. Further inclusion and exclusion criteria are shown in Table 2. Of 35 applicants, 20 experts qualified according to these criteria. 70 % of the examiners were male and 30 % were female, with an average age of 40.2 ± 13.9 years. The study complies with ICH-GCP standards and did not require the approval of an ethics committee. The qualifications of the participants ranged from at least a bachelor's degree in ophthalmic optics/optometry up to the highest qualification of PhD or MD. The survey on personal and professional data revealed that 75 % of the participants worked in the field of contact lenses, 10 % in optometry with a focus on the anterior segment of the eye, and 5 % in research. The average patient frequency was 12 ± 14 patients per week.

Table 1: Characteristics of the diagnostic images for online classification; n = 25

	Frequency
Eye under study	57 % right; 43 % left
Gender	62 % male; 38 % female
Ethnic origin	52 % Caucasian; 24 % Oriental; 19 % Asian; 5 % African
Dry eye	52 % without symptoms
Degree of redness	29 % grade 0; 43 % grade 1; 14 % grade 2; 10 % grade 3; 5 % grade 4

Table 2: Inclusion and exclusion criteria

| Inclusion criteria	Exclusion criteria
Paper-signed or digital declaration of consent	Insufficient theoretical/practical knowledge to classify the degree of redness of the conjunctiva
Good knowledge of the German and English languages	Unfinished classification of findings on the website
Completed professional training in the required area of expertise	40 % of a subject's findings are outside the results achieved by the overall collective of examiners
Technical equipment with internet connection and possibility of displaying dynamic web content	

Subjective classification

A special website was created in cooperation with the device manufacturer (Oculus Optikgeräte GmbH) to collect the diagnostic images for classification. A personal access key and an information letter were provided to access the website. The classification could only be started after the declaration of consent was signed and the agreement regarding the data protection declaration was given. The subjective classification of the diagnostic images was carried out as part of a prospective transversal study. The 25 diagnostic images were randomised on the website. The different images had to be classified separately three times for limbal and three times for bulbar redness. In total, each expert classified 150 images. The respective images of the JENVIS classification key served as reference images. These can be divided into five degrees of redness severity, whereby grade 0 stands for no redness and grade 4 for very severe redness (**Figure 1**).

The severity of eye redness was evaluated on the basis of 101 images displaying a progression in redness severity.[7] The original conjunctival images were then classified using software-developed sliders. Three different sliders were used to change the reference images so that the degree of redness matched that of the original image. The respective degree of redness could be seen in the field on the bottom part of the screen. Each slider changed the reference redness for the temporal, nasal and total redness of the bulbar conjunctiva respectively (**Figure 2**). To ensure that all examiners had the same amount of time and knowledge about the study content on the webpage, the study website was closed after 24 hours. If the classification run was not completed within the specified 24-hour period, it had to be repeated from the beginning.

Objective classification

The objective classifications were performed using the R-Scan (version 6.07r17) of the keratograph 5M (K5M). The first step was to take an image of the entire eye using the K5M. It was then possible to separate between different wavelengths from the image, allowing red, green and blue colour components to be analysed. The bulbar conjunctiva was separated by detecting the iris, eyelids and eyelashes. After automatic masking of the scleral blood vessels, the red colour components were detected and classified according to their area and quantity (**Figure 3**).[14] Studies have shown that
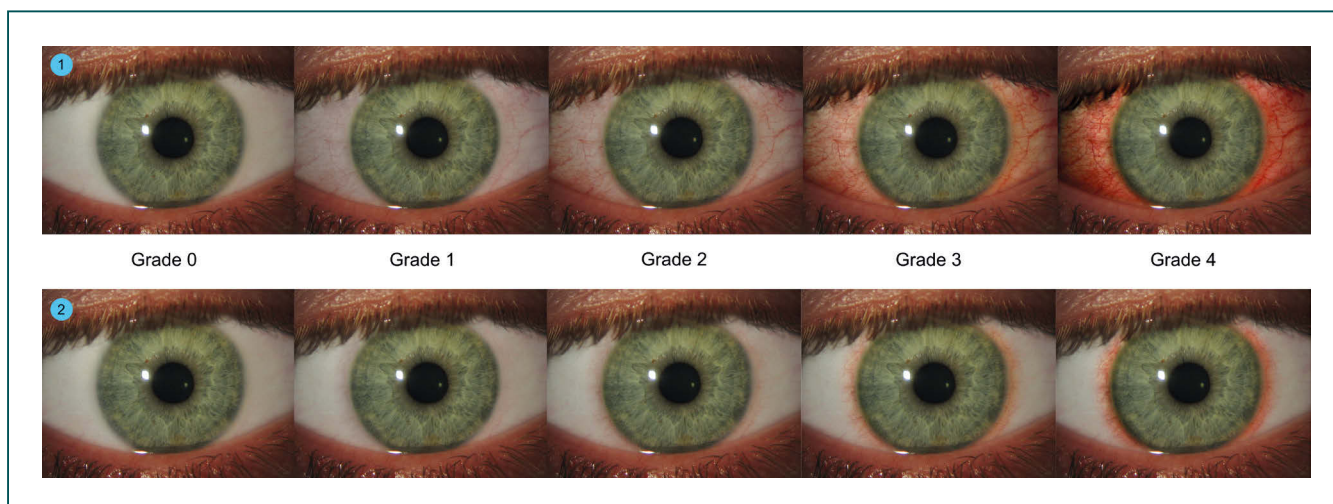
Figure 1: Schematic representation of the JENVIS grading scale divided into (1) bulbar injections and (2) limbal injections
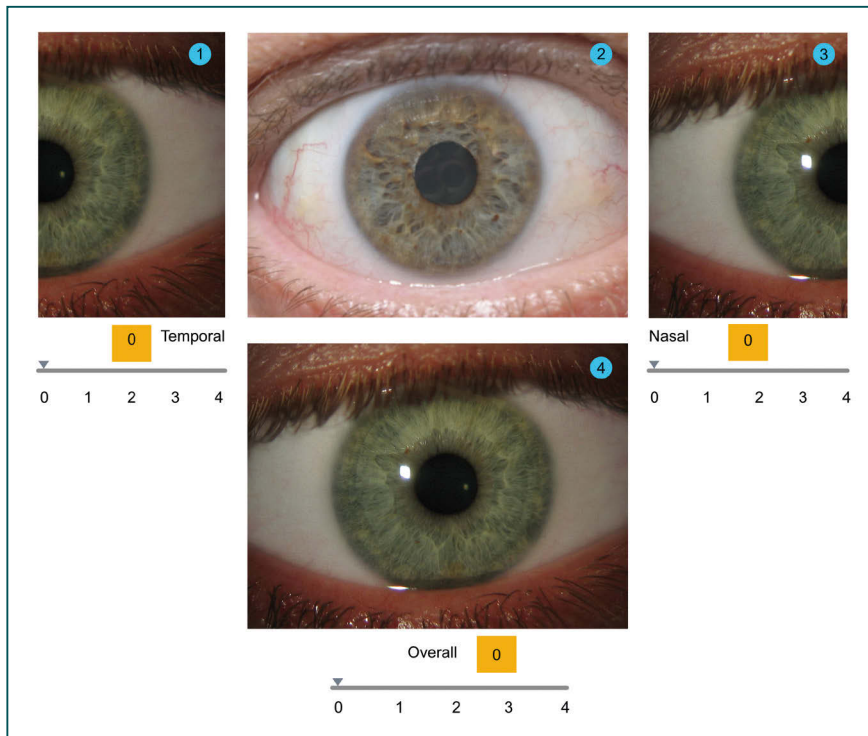
Figure 2: Overview of the classification software – (1) assessment of the temporal eye area, (2) original image, (3) assessment of the nasal eye area, (4) assessment of the entire eye

Figure 3: R-scan procedure to determine conjunctival redness (from left): original image, exposed area of the bulbar conjunctiva and filter procedure to detect the blood vessels[14]

the K5M offers higher reproducibility and better repeatability between different examiners than is possible with subjectively interpreted grading scales thanks to precise focus adjustment and constant lighting conditions during the measurements.[16,17]

Statistical analysis

The subjective results were compared with the analysed degrees of redness of the objective results using the R-Scan (Keratograph 5M, Oculus, Germany). Statistical data analysis was then performed using the version 20 of the SPSS software (IBM, USA), Microsoft Excel (Microsoft, USA) and MedCalc statistical software 11 (MedCalc Software, Belgium). To analyse the variability of the results of the two classification groups, the arithmetic mean per examiner was calculated from three repeated measurements and sorted chronologically. In addition, the standard deviation of the random sample and the 95 % confidence interval were determined. To compare the mean values between the subjective and the objective

classifications, the individual results were grouped by the overall grade (score 0-4, steps of 1). The overall grades were determined by averaging the classification grades from the objective and the subjective classifications. The difference between the results was calculated by subtracting the objective classification from the subjective classification. The normal distribution was tested using the Kolmogorov-Smirnov test with Liffifors constraints. The normal distribution was assumed with a statistical p-value greater than 0.05. The Shapiro-Wilk test was used to check the test results. If a difference in the assumption of a normal distribution between the two test procedures (Kolmogorov-Smirnov and Shapiro-Wilk) was observed, the results of the Kolmogrov-Smirnov adjustment test were used for large samples (> 50 measurements per variable) in accordance with Brosius' recommendation, otherwise those of the Shapiro-Wilk test were used. The intraclass correlation coefficient (ICC) was determined to demonstrate the quality of the measurement methodology.[15] The double t-test with a significance level of 5 % was used for the statistical comparison of the mean values. No distinction

Table 3: Descriptive analysis of the objectively recorded findings using R-Scan (n = 25)

		Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
Total redness						
	Percentage	24 %	48 %	14 %	14 %	—
	R-Scan	0.4 ± 0.0	1.0 ± 0.1	2.1 ± 0.1	3.1 ± 1.6	—
Bulbar redness						
Temporal	Percentage	19 %	52 %	19 %	5 %	5 %
	R-Scan	0.5 ± 0.0	1.0 ± 0.1	2.1 ± 0.2	3.5 ± 0.1	3.7 ± 0.0
Nasal	Percentage	19 %	52 %	14 %	14 %	—
	R-Scan	0.4 ± 0.0	1.0 ± 0.1	2.1 ± 0.2	2.9 ± 0.2	—
Limbal redness						
Temporal	Percentage	43 %	48 %	5 %	—	5 %
	R-Scan	0.4 ± 0.0	1.0 ± 0.2	2.5 ± 0.2	—	3.8 ± 0.2
Nasal	Percentage	38 %	43 %	14 %	5 %	0 %
	R-Scan	0.3 ± 0.1	0.9 ± 0.1	2.1 ± 0.2	2.6 ± 0.7	—

was made between the groups in the basic comparison of the mean values. The post-hoc analysis was carried out within the classification groups. The null hypothesis assumed that the mean values were equal.

Results

Table 3 lists the objective findings of bulbar redness. The differences between the subjective classification and the objective measurements are shown in Table 4.

The results of the data analysis to test the agreement in redness-grade determination between the subjective classification and the objective measurements as well as the spread can be found in Tables 5 and 6.

Discussion

The comparison of the subjective and objective classifications regarding their agreement per test subject was determined using the ICC and Cronbach's α and documented in Table 5. In this context, the overall redness shows the best agreement between the two results, since the lower limit of the 95 % confidence interval (CI) is 0.7. The lowest agreement or the highest variability between the results in the comparison of subjective and objective classification was achieved by the nasal-redness ratings, both bulbar and limbal, with a 95 % CI of the ICC of 0.352 to 0.916 for bulbar nasal redness and a 95 % CI of the ICC of 0.171 to 0.915 for limbal nasal redness. The classifications of the temporal side of the eye showed medium to very good agreement. Similarly, a direct compar-

ison of the two types of classification showed the smallest differences in the assessment of overall redness.

The contingency table (Table 4) showed a high degree of agreement between the two types of classification in grades 0 and 1. Within the two grades, most test subjects (over 50 %) were assessed with the same grade. For example, the same patient was assessed as grade 1 based on redness both by the R-scan and on average subjectively. As the grade increases, the deviations between the two classifications also increase, meaning that the majority of the test subjects were assessed differently from grade 2 onwards. Differences in the subjective and objective assessment were particularly evident at grade 3. One reason for this is the type of redness of the affected eye. The R-Scan software, only considered blood vessels to analyse the degree of redness, whereas in the subjective classification, not only the volume of blood in the vessels was taken into account, but the entire subjectively evaluated redness of the eye, as well as scleral and episcleral vessels. The greatest difference between the two types of classification was found in eyes exhibiting hyposphagma. For this reason, further development of the software should focus on the type of redness in the eye that is actually determined. If a classification of the redness of the eye is assumed, areas with haemorrhage should also be analysed, as these induce increased redness. However, if a classification of conjunctival injection or conjunctival hypo- or hyperaemia is the goal, then only the blood vessels present and their blood volume should be analysed, as the current state of the software does.

The comparison of the mean values revealed significant differences between subjective and objective classification for total redness, nasal bulbar redness as well as temporal and nasal limbal redness. The post-hoc analysis showed that these

Table 4: *Statistical analysis of the results of the total redness from the subjective and objective classifications of the diagnostic images (n = 25)*

Total redness	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
Subjective	0.3 ± 0.2	1.1 ± 0.3	2.0 ± 0.3	3.3 ± 0.3	—
Objective	0.4 ± 0.0	1.0 ± 0.1	1.9 ± 0.1	2.5 ± 1.2	—
Difference	−0.1 ± 0.1	0.2 ± 0.2	0.1 ± 0.3	0.8 ± 1.4	—
p-value (total) = 0.041					
p-value (single)	0.015	0.04	0.351	0.164	—

Table 5: *Agreement and correlation analysis between the objective and the subjective classifications (n = 25)*

	Cronbach's α	ICC	95 % KI (ICC)	p-value	Pearson correlation coefficient	p-value
Total redness						
	0.855	0.845	0.707; 0.917	0.000	0.762	0.000
Bulbar redness						
Temporal	0.836	0.720	0.585; 0.930	0.000	0.720	0.000
Nasal	0.830	0.778	0.352; 0.916	0.000	0.754	0.000
Limbal redness						
Temporal	0.850	0.834	0.590; 0.933	0.000	0.718	0.000
Nasal	0.839	0.758	0.171; 0.915	0.000	0.772	0.000

Table 6: *Analysis of the spread within the redness grades. grouped by classification type and evaluation area (n = 25)*

	Classification		Grade 0	Grade 1	Grade 2	Grade 3	Grade 4	Total
Total redness		objective	0.00	0.07	0.07	0.39	—	0.13
		subjective	0.70	0.27	0.18	0.11	—	0.32
Bulbar redness	temporal	objective	0.00	0.06	0.09	0.12	0.05	0.06
		subjective	0.35	0.28	0.19	0.15	0.06	0.21
	nasal	objective	0.26	0.04	0.08	0.15	—	0.13
		subjective	0.61	0.31	0.19	0.12	—	0.31
Limbal redness	temporal	objective	0.00	0.08	0.17	0.04	0.00	0.06
		subjective	0.75	0.28	0.17	0.15	0.17	0.30
	nasal	objective	0.13	0.12	0.15	0.17	—	0.14
		subjective	0.77	0.30	0.18	0.15	—	0.35
	objective (total)		0.07	0.06	0.10	0.18	—	0.11
	subjective (total)		0.64	0.29	0.18	0.13	—	0.31

significant differences were determined in grades 0 and 1; in the case of nasal limbal redness significant differences were only determined for grade 1. These differences could be attributed to the composition of the sample. The different sample sizes within the redness grades clearly changed the respective effect size for the detection of a significant difference. However, after preselecting the sample size, the effect size was the only variable. At higher levels of redness, a more than significant difference had to exist in order to detect a statistically significant difference, whereas at lower levels of redness, a small effect was sufficient. The analysis of the spread of repeated measurements per subject showed that the spreading width of objective classification increased with increasing redness but was significantly smaller than the spread of subjective classification (Table 6). In this context, the mean objective dispersion was 11 % and the mean subjective dispersion 31 %. In contrast to the objective classification, the spread of the subjective classification decreased with increasing redness.

The measurement results were influenced, for example, by the selection of possible investigators. This was subject to a selection bias, as all the experts contacted came exclusively from Europe. Accordingly, the practical experience was mainly limited to Caucasian eyes. The classification of Asian, Oriental and African eyes was therefore carried out with a lower level of experience. However, the influence of the selection bias was low, as the spread within the different ethnic groups varied little. Furthermore, to avoid the influence of the learning effect or the control of the results, the participants in the online study were not able to change their classification assessment. Based on this, it was possible that participants confirmed their classification inadvertently or too quickly.

Conclusion

The software (R-Scan, K5M) primarily assessed the number and redness of the conjunctival blood vessels, whereas the subjective assessment was based on the overall redness perceived by an examiner, including scleral and episcleral injections of the eye. The spread of the results decreased for both the objective classification and the subjective classification with increasing redness, with the objective classification showing an almost 3-times greater spread on average. Thus, automated conjunctival classifications were more accurate than subjective classifications at lower degrees of redness and comparable at higher degrees.

Conflict of interest

The author declares that there is no conflict of interests regarding the methods and devices mentioned in the article.

Corresponding author

Dr. Sebastian Marx

E-Mail:
s.marx@jenvis-research.com

References

1 Burk, A., Burk, R., Kaden, R. (2019). Differentialdiagnose in der Augenheilkunde. Vom Befund zur Diagnose. 1st ed. Heidelberg, Kaden Verlag, Heidelberg.

2 Sachsenweger, M. (1994). Augenheilkunde. Stuttgart: Hippokrates Verlag.

3 Papas, E. (1998). On the Relationship Between Soft Contact Lens Oxygen Transmissibility and Induced Limbal Hyperaemia. Exp. Eye Res., 67, 125–31.

4 Efron, N. (2023). Contact Lens Practice 4th Edition, Elsevier, Inc.

5 Kunert, K., Sickenberger, W., Brewitt, H. A. (2016). Trockenes Auge. Anatomie, Physiologie, Pathophysiologie, Diagnostik und Therapie des Sicca-Syndroms. Kaden Verlag, Heidelberg.

6 Terry, R. L., Schnider, C. M., Holden, B. A., Cornish, R., Grant, T., Sweeney, D., La Hood, D., Back, (1993). CCLRU Standards for Success of Daily and Extended Wear Contact Lenses. Optom. Vis. Sci., 70, 234–243.

7 Sickenberger, W., Oehring, D. (2012). Validation of a Novel Morphing Software to Classify Different Slit Lamp Findings. Cont. Lens Anterior Eye, 35, e20-e21.

8 Müller, A.(1889). Brillengläser Und Hornhautlinsen. Diss., Kiel, Druck von L. Handorff.

9 Murphy, P. J., Lau, J. S. C., Sim, M. M. L., Woods, R. L. (2007). How Red Is a White Eye? Clinical Grading of Normal Conjunctival Hyperaemia. Eye (Lond.), 21, 633–638.

10 Pult, H., Murphy, P. J., Purslow, C., Nyman, J., Woods, R. L. (2008). Limbal and Bulbar Hyperaemia in Normal Eyes. Ophthalmic Physiol. Opt., 28, 13–20.

11 Fieguth, P., Simpson, T. (2002). Automated Measurement of Bulbar Redness. Invest. Ophthalmol. Vis. Sci., 43, 340–347.

12 Schulze, M. M., Hutchings, N., Simpson, T. L. (2008). The Use of Fractal Analysis and Photometry to Estimate the Accuracy of Bulbar Redness Grading Scales. Invest. Ophthalmol. Vis Sci., 49, 1398–406.

13 Schulze, M. M., Hutchings, N., Simpson, T. L. (2011). Grading Bulbar Redness Using Cross-Calibrated Clinical Grading Scales. Invest. Ophthalmol. Vis Sci., 52, 5812–5817.

14 Jerchel, N. (2011). Objektive Klassifikation und Dokumentation der bulbären Bindehautrötung mittels Videokeratographie. EAH Jena.

15 Fleiss, J. L., Cohen, J. (1973). The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. Educational and Psychological Measurement, 33, 613–619.

16 Wu, S., Hong, J., Tian, L., et al. (2015). Assessment of Bulbar Redness with a Newly Developed Keratograph. Optom. Vis. Sci., 92, 892–899.

17 Müller, J. (2019). Ermittlung des Rötungsgrades der Bindehaut von weichen Monatslinsenträgern durch Umstellung auf eine weiche Silikon-Hydrogel-Tageslinse. EAH Jena.